



Data Architecture & Engineering

QUAN NGO

(+84) 962269422 | katoo2706@gmail.com

<https://katoengineer.com/>

<https://www.linkedin.com/in/katoo2706/>

Hanoi, Vietnam

Work Experience

"Experience delivering end-to-end data solutions across product and outsourcing environments, with a focus on data architecture, scalable pipelines, cloud infrastructure, and cross-functional team leadership"

FPT SOFTWARE (Software company / Outsourcing)

Apr 2024 - **Now**

Highlights: Lead Data Engineer at GHC DHM (a business unit at FPT software), collaborating with the consulting team to design data-driven solutions for bidding processes.

Project:

1. Healthcare Enterprise Data Platform / Healthcare client in US

Role: Senior Data Engineer.

- Work in global Enterprise data team in Healthcare
- Design & implement Standard data warehouse & data pipeline.
- Technical leader of engineers team (teamsize: 5)
- CICD design & development for data pipeline, dbt cloud, Snowflake schema change.

Tech stacks: Snowflake, AWS cloud, DBT cloud, Pyspark. / Python + SQL

2. Data Lakehouse platform / Healthcare client in Vietnam

Role: Technical Leader– Data Architect

- Design and manage Lakehouse data-platform using micro-services architecture on Kubernetes
- Technical leader of Engineers team including DE, DA, BE (teamsize: 20)
- Design, develop realtime data pipelines to sync healthcare data from multiple sources.
- Design & implement the healthcare event-driven Application using Java SpringBoot & Flink.

Tech stacks: Kubernetes, Trino, Jenkins, DBT, Spark, Kafka, Flink, Starrock, Apache Airflow, Apache Ranger, Spring Boot, Superset, Grafana/Prometheus, ELK.

BREADSTACK TECHNOLOGIES (Ecommerce & Software – Canada / Product)

Oct 2022 – Apr 2024

Role: Data Engineer, Data Architect.

General responsibilities:

- Design and manage all data architecture for these projects:
 - Embedding analytics dashboard on software (Embed etabase dashboard on web app)
 - Sales & Marketing Data warehouse project (Datawarehouse platform build on GCP using Clickhosue, Airflow, Polars, PowerBI).
 - HR project Data warehouse project (Data warehouse for HR team to integrate with 1Office & Humi for employees management)
- Design data models and integrate all below data sources:
 - Google Analytics, GSC, CRM, ERP system, Klaviyo, Social Platforms (Facebook / Instagram / Twitter / LinkedIn).
 - Company's CRM softwares: Breadstack, Chatso.

- Task management: Jira, Trello.

○

DATA ENGINEER MENTOR – FUNiX Technologies School (Education - Vietnam)

Jul 2023 - Now

General responsibilities: Mentored at FUNiX School, providing practical guidance in data engineering & analytics.

TIEN PHONG BANK (Banking & Finance – Vietnam / Product)

Jul 2022 – Oct 2022

Role: Analytics Engineer (at ALM department)

General responsibilities:

- Build Python module to auto-cleanse data, build reports, and ETL data automatically
- ALCO report / GAP report.
- Daily FTP data management.
- Coordinating implementation of liquidity management, and optimizing cash flow on the balance sheet scale.
- Make management reports as required and assigned.
- Build automatic EDA with Python notebook, transform old report to BI visualization,

Techstacks: Python · PowerBI · SQL · Qlikview

HONG NGOC GROUP (HealthCare – Vietnam / Product)

Nov 2019 - May 2022

Role: Data Analyst

General responsibilities:

- Manage Services' cost components, Services' pricing.
- Work with BE team to improve the data system.
- Building report dashboard, make ad-hoc by MS. Excel & Power BI + SQL
- Cleansing data & Analyze Revenue, Profit, ROS and KPI of branches.
- Improve and build the whole new Report dashboard, and improve data processing method of the team.

Techstacks: Microsoft Excel, SQL, PowerBI, Python

Familiar with

“Passionate about building scalable architecture, data pipelines, design & managing infrastructure, and ensuring data quality across modern platforms.”

-
- **Database:** Relational databases (**PostgreSQL, MySQL, AWS RDS, Cloud SQL**), non-relational databases (MongoDB, Cassandra, DynamoDB).
 - **Programming Languages:** Python (Fast API, Flask, Django), Java (SpringBoot), Scala, SQL.
 - **Data Processing:** Batching (using Spark, dbt) & streaming data pipeline (Kafka, Apache Flink, Spark streaming).
 - **Pipeline Orchestration:** Apache Airflow, Azure Data Factory, and Databricks.
 - **Data Warehousing:** Snowflake, BigQuery, Redshift, and ClickHouse, Trino, Starrock.
 - **Lakehouse & Data Lakes:** Designed, implement Lakehouse architecture using Trino, Hive Metastore, Iceberg, and object storage (Google Cloud Storage, AWS S3, Azure Data Lake, and MinIO).
 - **Cloud Services:** Extensive experience with AWS, Databricks, Google Cloud, Microsoft Azure, and Cloudflare.
 - **DevOps:** Skilled in **Kubernetes (GKE), Terraform, Jenkins, CI/CD pipelines**, and microservices architecture.
 - **Data Governance & Quality:** Experience with Apache Ranger, OpenMetadata, ydata_profiling, and Great Expectations.
 - **Data Visualization:** Skilled with enterprise tools like Microsoft Power BI and Looker, as well as open-source platforms such as Superset, Metabase, and Plotly.
 - **Soft Skills:** Strong problem-solving abilities, **technical leadership**, and mentoring experience.

Achievements

"I have over 5 years of experience in the Data industry and am on my way to becoming the best Data Architecture."

Certified Data Management Professional Certification

Sep 2025

DAMA International

Desc: Certified Data Management Professional (CDMP) is a globally recognized Data Management Certification program run by DAMA International.

Databricks Certified Data Engineer Associate

July 2025

Desc: Specializes in designing, deploying, and maintaining foundational data engineering solutions using the Databricks Lakehouse Platform. Proficient in building ETL pipelines with Spark SQL and Python, managing multi-hop architectures, and operationalizing queries and dashboards.

Best Performer - FPT Software

January 2025

Desc: Specializes in designing, deploying, and managing scalable data solutions on the AWS platform, with proficiency in data lakes, data warehouses, ETL processes, and leveraging services such as S3, Redshift, Glue, and EC2 for efficient data handling and analysis.

AWS Certified Data Engineer

June 2024

Desc: Specializes in designing, deploying, and managing scalable data solutions on the AWS platform, with proficiency in data lakes, data warehouses, ETL processes, and leveraging services such as S3, Redshift, Glue, and EC2 for efficient data handling and analysis.

IBM Data Engineering Professional Certificate

June 2023

Desc: Python, NoSQL databases, relational databases (RDBMS), SQL, as well as familiarity with Hadoop and Spark, and creating data warehouses utilizing business intelligence tools.

Nationals Economics University – Bachelor's degree

June 2015 - June 2019

Programs: Advanced Education Program / Economics Investment

Appendix

"For me, certificates are not an end in themselves. As far as I am concerned, it's about the progress to achieve all of that."

- Apache Spark Streaming 3 with Scala	Rock the JVM
- Rest APIs with Flask and Python	Udemy
- The Complete dbt (Data Build Tool) Bootcamp Zero to Hero	Udemy
- Apache Spark 3 & Big Data Essentials in Scala	Rock the JVM
- Azure Databricks & Spark For Data Engineers (PySpark / SQL)	Udemy
- Data Engineering and Machine Learning using Spark	Coursera
- Introduction to Big Data with Spark and Hadoop	Coursera
- Modernizing Data Lakes and Data Warehouses with Google Cloud	Google Cloud Training Online
- ETL and Data Pipelines with Shell, Airflow and Kafka	Coursera
- AWS Cloud Practitioner Essentials	AWS Training Online
- Databases and SQL for Data Science with Python (with Honors)	Coursera
- Python for Data Science, AI & Development	Coursera

Reference: <https://www.linkedin.com/in/katoo2706/details/certifications/>